PE - 16

THE ESTIMATION OF INQUIRY PERFORMANCE TEST ITEMS OF HIGH SCHOOL PHYSICS SUBJECT WITH QUEST PROGRAM

By: Supahar
Physics Education Department, Mathematics and Natural Science Faculty, Yogyakarta State University
e-mail: pahar.fis@gmail.com, supahar@uny.ac.id

Abstract

This paper is aimed to determine the results of the parameter estimation of inquiry performance test item of high school physics subjects with polytomous pattern using the QUEST program. The estimation of the item test based on the item response theory (IRT) by applying the model of Partial Credit Model (PCM) on the student's response of inquiry performance test. The estimation included fit model the PCM, and the estimation of test items difficulties level. Based on the estimation of test items parameters on the IRT according to the model of PCM known that the items of inquiry performance test have an INFIT scale of MNSQ between 0.77 to 1.30 thus meet the fit criteria with PCM models, and test item difficulties fulfilling the criteria of a good test item because of the test item difficulties in the range of $-2 < b < +2$.

Keywords : *parameters of the test item , Partial credit model (PCM) , inquiry performance, high school physics subjects*

INTRODUCTION

Operational definition on science, which includes physics, has been stated by experts. Moh. Amien (1987: 4) states that science is a set of knowledge obtained using methods based on observation. Another approach to define science (Kane & Sternheim,1988: xii) is by asking what scientists are thinking. Scientists try to understand basic rules or laws which determine the operation of the natural world. According to Carin & Sund (1989: 4) science is a system to discover the universe through data collection by observations and controlled experiments. There are three criteria which must be met by a theory in science. The three criteria are (1) ability to explain phenomenon which have been observed or have occurred, (2) ability to predict phenomenon which will occur, and (3) can be tested with similar experiments. Based on experts' opinions, science definition includes three main elements, i.e. science as a process or method, science as a product in the form of facts, and science as human attitude in the form a particular belief.

Collette & Chiappetta (1994: 30), Chiappetta & Koballa (2010: 104-114) state that science is essentially a body of knowledge resulted from inquiries, a way of thinking to understand nature, and a way of investigating to discover natural phenomenon. Referring to this statement, scientists' views stated above aren't wrong. Each is just one of the natures of science in that statement.

So, the natures of science, including physics, can be stated as follows. (1) Science as a product to replace the statement of science as a body of knowledge. (2) Science as an attitude to replace the statement of science as a way of thinking, and (3) science as a process to replace the statement of science as a way of investigating. Considering physics is a part of science, the view on the natures of science can be compared with the natures of physics. Therefore, the nature of physics are: (1) physics as a product aspect or a body of knowledge, (2) physics as an attitude aspect or a way of thinking, and (3) physics as a process aspect or a way of investigating.

Physics in high school provides learning experience to understand concepts, and train inquiring ability using science process skills related to natural phenomenon. Some of the demands in the formulation of basic competence (KD) in national curriculum content standard for high schools are: (1) experimenting, and (2) understanding measurement principles, (3) and measuring physical quantity directly and indirectly carefully, conscientiously, and objectively. Therefore to reach basic competence of High School physics, inquiring ability is required.

National Science Education Standards in Wenning (2005: 3) defines inquiry for students as student activities to develop knowledge and understanding on scientific ideas, and understanding how scientists study the natural world. Moyer, *et.al*. (2007:5) defines inquiry and scientific processes which reflect scientific methods. Bass, *et.al*. (2009: 17) defines inquiry as scientists' way to investigate. Moh Amien (1987: 126) states that inquiry includes discovery. Therefore, high school students involved in inquiry must learn scientific methods which include planning, implementation, and reporting, including communicating investigation results.

Chiappetta & Koballa (2010: 130) state that inquiring ability includes science process skill strategies, asking questions, discrepant events, inductive activities, deductive activities, gathering information, and problems solving. The skill in making models in the form of flow charts or making mathematic models as well as the ability to communicate observation or experiment results are also parts of science process skills.

Science process skills as an inquiring ability according to education experts include the following aspects and sub-aspects. Scientific processes according to Bryce, *et.al.* (1990:1-4) consist of: (a) basic skills aspect which consists of: observation skill, data recording skill, measuring skill, manipulating skill, procedural skill, and skill of following instructions and (b) process skills aspect which consists of: inference skill and skill of selecting procedures. Scientific processes according to Rezba, *et.al.* (1995) consist of: (a) basic science process skills, which include observing, communicating, classifying, measuring, inferring, predicting, and (b) integrated science process skills which consist of: identifying variables, arranging data into tables, making graphs, illustrating relation between two variables, collecting and processing data, analyzing investigation result, making hypothesis, defining variables operationally, designing investigations, and experimenting.

Furthermore, scientific processes according to t Bass, *et. al*. (2009: 30) consist of the following sub-aspects: (a) observing, (b) classifying, (c) inferring, (d) measuring, (e) communicating, (f) predicting, (g) hypothesis, and (h) experiment: controlled investigation. The same thing but in different order is stated by Carin, *et. al*. (2001: 44), that scientific processes consist of the following sub-aspects: observing, measuring, classifying, inferring, hypothesis, controlled investigation, predicting, explaining, and communicating results. Scientific processes according to Bambang Subali (2009: 583-586) consist of the following sub-aspects: (a) basic skills which consist these sub-aspects: observation skill, data/information recording skill, skill in following instructions, measuring skill, movement manipulation skill, and procedure/technique/use of equipment skill, (b) processing skills consist of these sub-aspects: predicting skill, inferring skill, and procedure selecting skill, and (c) Investigating skills consist of these sub-aspects: investigation designing skill, investigating skill, investigation result reporting skill.

Based on experts' opinions above, it can be said that scientific processes as inquiring ability consist of: (1) investigation planning, (2) investigation implementation, and (3) investigation results reporting. Investigation planning sub-aspect indicators consist of these sub-aspects: designing investigation and selecting procedure. Investigation implementation sub-aspect indicators consist of these sub-aspects: observing, recording data/information, following instructions, measuring, manipulating movements, implementing procedures/techniques/use of equipments, and investigating. Reporting sub-aspect indicators consist of these sub-aspects: predicting, inferring, reporting investigation results, and explaining investigation results. In short, inquiry through scientific processes is illustrated in figure 1.

Assessment system in schools has been enacted in the Regulation of Minister of National Education Number 20 of Year 2007. Assessment format can be written tests, practice tests, and individual or group assignments. Glencoe science (t.t.: 3) divides education assessment features into two, i.e. traditional assessment features and performance assessment features. Traditional assessment is testing with paper and pencil test, such as multiple choices, right-wrong, and matching. Another form of paper and pencil test is asking testees to write their responses, such as essay items, whether short answer essays or open essays. Performance assessment shows how good students use one or more literature elements. Traditional assessment and performance assessment should be even. Sometimes traditional assessment can be used first to ensure that students have had accurate information before performance assessment is performed. Moreover, sometimes, performance assessment can be used first as strategy to involve students in learning.

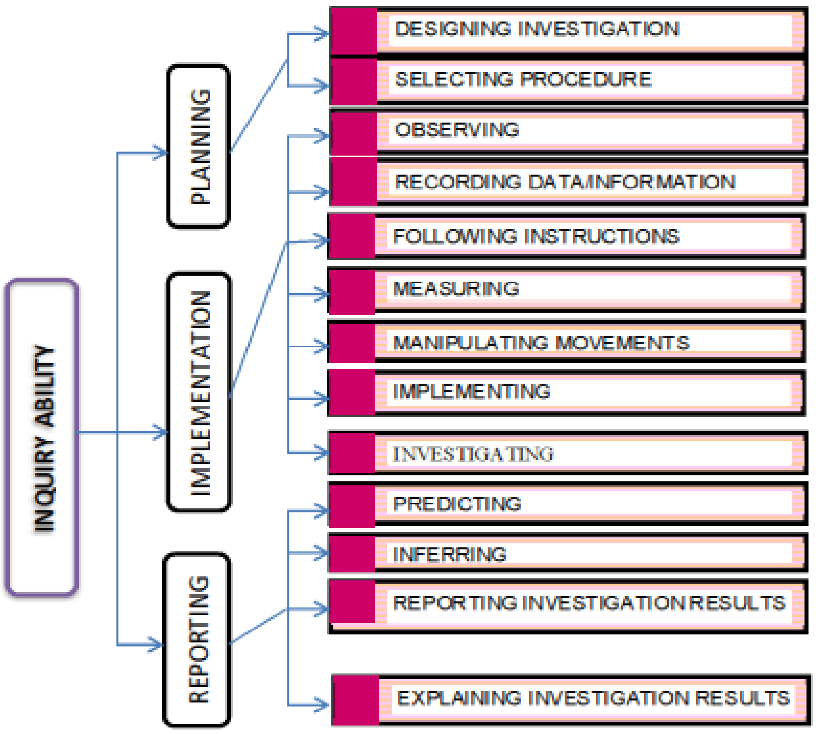

Figure 1. Inquiry Ability Indicators

Performance assessment is ideally performed using direct observation method which can be used as a benchmark. Other alternative methods as replacements of direct observation are: notebooks, computer simulation, and paper and pencil test (Ruiz-Primo & Shavelson: 1996: 1047-1050). Performance assessment by paper and pencil test can be used to reveal students' ability in scientific processes and inquiry procedures (Bass *et. al*.: 2009: 173). Performance assessment is performed in certain conditions which are imitations of real conditions. Testees are asked to show their ability in performing a task. Students' responses/answers are assessed and compared with established criteria in the rubric.

The result of performance assessment in the field of science in program for International Student Assessment (PISA), show that average performance score in the field of science of Indonesian students is in the 50th rank of 57 countries (Chiappetta & Koballa, 2010: 22-24). It shows that the performance ability in the field of science in general and physics in particular of Indonesian students is still low compared to other countries. Performance assessment by inquiry ability test is expected to measure the basis of inquiry knowledge, ability, and skill in high school physics, which include planning, implementation, and reporting abilities to give a more comprehensive illustration on students' achievements.

Instrument of performance assessment of high school physics inquiry ability has been developed by Supahar (2013: 235-244). The main problem which will be revealed in this paper is the characteristics of the instrument of performance assessment of each aspect of high school physics inquiry ability, which include planning, implementation, and reporting abilities? In accordance with the formulation of the problem which will be solved, the purpose of this paper is to discover the characteristics of the instrument of performance assessment of high school physics inquiry ability which includes, planning, implementation, and reporting abilities.

DISCUSSION

Question item analysis is an activity which must be done during the development of test instruments to improve the quality of questions which have been made. This activity is the process of collecting, summarizing, and using information from students' answers to make a decision on each assessment. The purpose of item analysis is to review every item to get quality questions before they are used. Furthermore, the purposes of question item analysis are also to help improve tests by revising or removing ineffective questions, and to discover diagnostic information on students, whether they have understood the materials taught or not. Quality questions are questions which can give accurate information in accordance with the purpose, including determining which student has or hasn't the material taught by the teacher.

Qualitative analysis includes content and construct validities consideration, while quantitative analysis includes item characteristic measurement. So, there are two ways which can be used in question item reviewing, i.e. qualitative and quantitative question review. Both techniques have strengths and weaknesses. Therefore, it's best to use both techniques.

Qualitative analysis of instrument of high school physics inquiry ability performance assessment which was developed showed that content validity was obtained by professional judgment from physicists, physics education experts, measurement experts, and practitioners. Quantitative analysis of instrument of high school physics inquiry ability performance assessment was based on empiric data from the relevant question items using Item Response Theory (IRT) approach.

Empiric data was obtained by measurement performed in Public High Schools in the Special Region of Yogyakarta from January to June 2013. Selection of schools as research location was based on National Exam ranking in 2012, and selected with purposive sampling. The selection Public High Schools was expected to represent large variation in students' characteristics, such as family background, culture, ethnicity, religion, social, and economy. While test subjects consisted of 2383 high school students of X class and XI-IPA class who were present during the test. XII-IPA class wasn't involved as research subject because they were facing National Exam 2013.

Empirical data collection technique was using written test. Data collection instrument was a set of high physics inquiry ability test. Written test set consisted of 93 items. Considering each meeting in class had 90 minutes, it's not possible for every student to finish all items with the available time allocation. Due to the consideration, the test design was divided into four test sets, i.e A, B, C, and D test sets. Every test sets was arranged by considering the representativeness of each inquiry ability aspect which would be measured. Every test set consisted of 30 items including 9 anchor items. Every respondent only had to do one test set provided under the supervision of the teacher. The advantage from this kind of test design includes minimizing cheating.

Test responses were polytomous data with four categories, i.e. category 1, 2, 3, and 4. Testee response data was analyzed using IRT according to Partial Credit Model (PCM) using QUEST program. *QUEST* program could be used to test: (1) goodness off it of items using PCM and test reliability determination, and (2) estimation of item difficulty level, Moreover, *QUEST* program could also be used to determine the estimation of students' ability (θ) which was stated in logit scale.

Goodness off it test on PCM was seeing the values of INFIT Mean of Square (Mean INFIT MNSQ) and its standard deviation. If average INFIT MNSQ was close to 1.0 with standard deviation close to 0.0 then all test items are fit with *PCM*. Test to determine the fitness of each item on PCM followed the principles established by Adam & Khoo (1996: 30 & 90), that an item is fit with a model if INFIT MNSQ value was in the range of 0.7 to 1.30. The range of INFIT MNSQ value limited the distribution of calibrated scores and was still in leptocurtic curve which reflected being in unity.

Item difficulty index testing used Hambleton & Swaminathan's (1985: 36) criteria that a test item is categorizes as very difficult if difficulty $>+2$ and categorized as very easy if $b<-2$. Therefore in the development of this test, items had the value of $-2\leq b\leq+2$.

Determination of test reliability criteria was based on Sumadi Suryabrata's (2000: 40) opinion, that test result to make decisions on individuals should use tests with reliability coefficient at least 0,90. The higher the reliability coefficient of a test, the smaller the possibility of mistakes when people decisions based on scores in the test.

Testing involved 2383 respondents from 13 Public High Schools in Special Region of Yogyakarta. Testing implementation was designed in such a way that students who sat side by side work on tests with different test set code. The time provided for the test was 2 class sessions (90 minutes). The test involved all physics teachers in each school as invigilators.

Test set consisted of 4 question packages. During the test, the four test packages with A, B, C, and D codes were tested simultaneously in each class and each student in a class received 1 test packaged to be done. Test sets with A, B, C, and D codes were tested on 627, 596, 604, and 556 respondents, respectively.

Students' response data was scored polytomous with 4 categories, which are category 1, 2, 3, and 4. Polytomous data of the 4 categories were analyzed in accordance with Partial Credit Model (PCM). The result of test fitness analysis seen from *INFIT* parameter for mean square (*MNSQ*) presented in table 1 showed that the instrument of high school physics inquiry ability performance assessment met fit statistic criteria according to *PCM*. Analysis result showed that 93 test items had INFIT MNSQ values between 0.77 and 1.3 which meant all test items were *fit* with *PCM (see figure 2)*. Reliability estimated based on item analysis showed alpha Cronbach index in high criteria, which was 0.93.

Table 1. Test statistic fit parameter on 0.50 chance level.

No.	Test parameter	Item estimation	Case estimation
1.	*INFIT MNSQ*	1.11±0.08	1.11±0.07
2.	*OUTFIT MNSQ*	1.11±0.08	1.11±0.10
3.	Average difficulty	0.0±0.25	
4.	Reliability	0.93	

Item specification in test sets with A, B, C, and D codes had difficulty index between -1.15 (item5) in easy category to +0.70 (item2) in medium category. Average difficulty was 0.0±0.25 in medium category. Distribution of difficulty of items of high school physics inquiry ability performance test is presented in figure 3. The difficulty index of average test item for each aspect of measured inquiry ability, planning, implementation, and reporting, respectively, are presented in table 2.

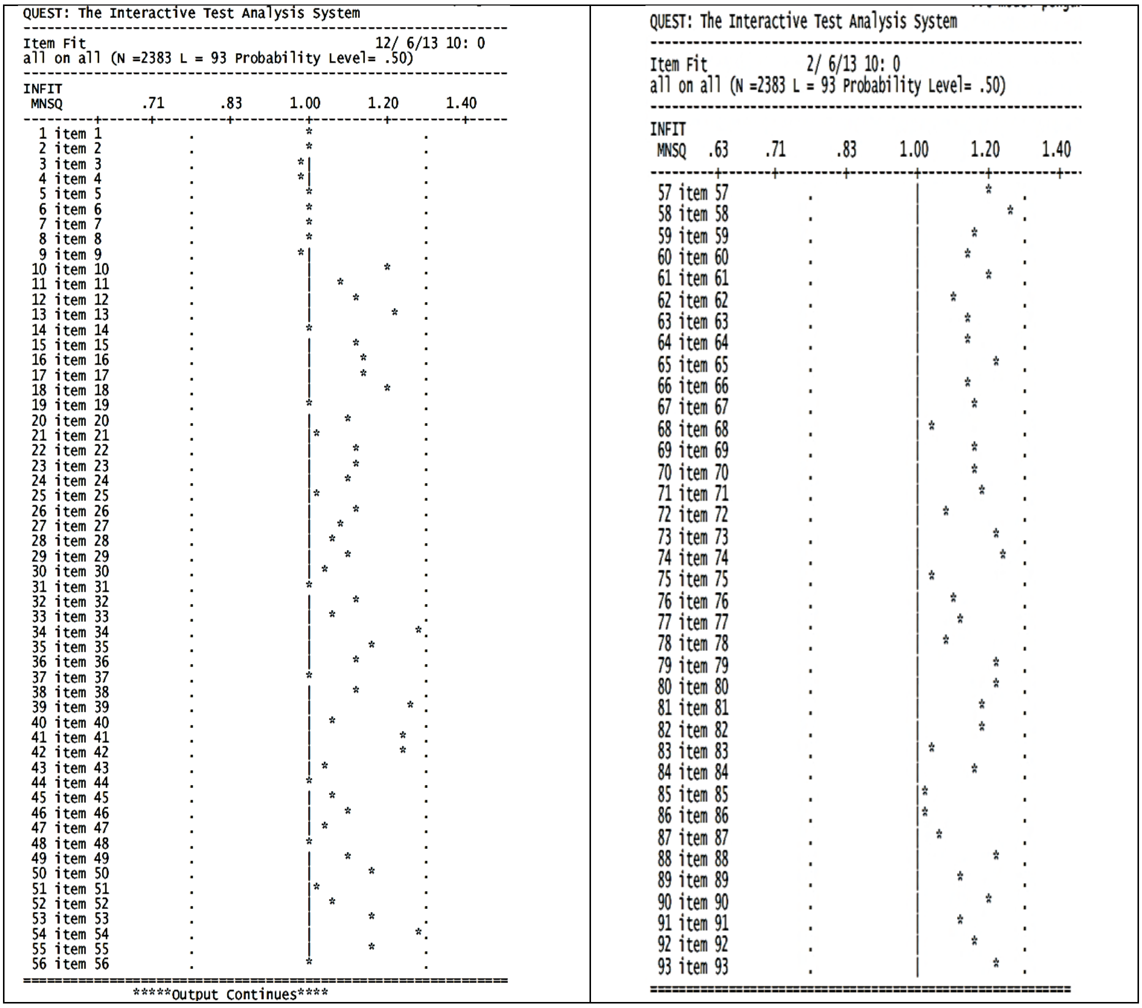

Figure 2. Items Fit with PCM

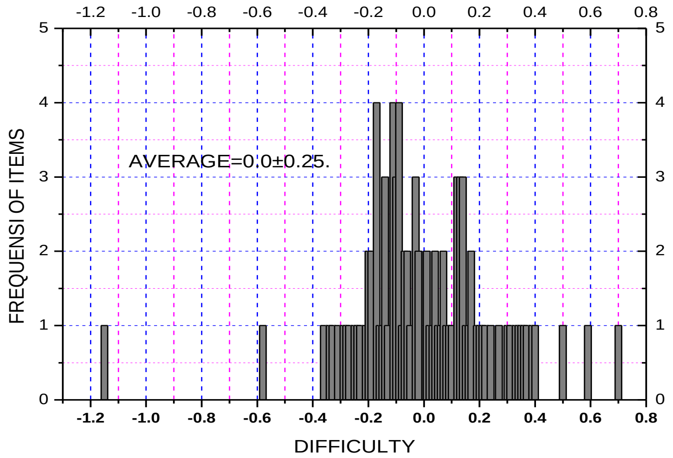

Figure 3. Distribution of difficulty index of high school physics inquiry ability test items

Table 2. Average difficulty of aspects and sub-aspects of high school physics inquiry ability

No.	INQUIRY ABILITY ASPECTS/SUB-ASPECTS	AVERAGE DIFFICULTY	DIFFICULTY INDEX CATEGORY			
			1	2	3	4
1.	PLANNING	0.095				
	a. Designing investigation	0.11	-1.08	0.53	1.44	-0.57
	b. Selecting procedure	0.03	-0.97	-0.01	1.66	-0.58
2.	IMPLEMENTATION	-0.01				
	a. Observing	-0.04	-1.39	0.49	1.17	-0.51
	b. Recording data/ information	0.04	-0.88	0.44	1.16	-0.56
	c. Following instructions	-0.20	-1.13	-0.59	1.38	-0.46
	d. Measuring	0.11	-0.95	0.76	1.01	-0.53
	e. Manipulating movements	-0.14	-1.11	-0.31	1.38	-0.50
	f. Implementing procedures/techniques/use of equipments	-0.09	-1.21	0.30	1.06	-0.50
	g. Investigating	0.16	-0.77	1.34	0.64	-0.55
3.	REPORTING	-0.07				
	a. Inferring	-0.11	-1.61	0.56	0.85	-0.33
	b. Predicting	-0.15	-1.19	-0.07	1.19	-0.53
	c. Reporting investigation results	0.02	-1.03	0.56	1.04	-0.54
	d. Communicating investigation results	0.12	-1.73	0.28	1.8	

CONCLUSION AND SUGGESTION

Conclusion

Based on the discussion, it was concluded that the instrument of inquiry ability performance assessment test developed on planning, implementation, and reporting abilities consisted of test sets A, B, C, and D, each of which had 30 question items with 9 anchor items as follows.

1. Inquiry ability performance assessment test instrument met content validity with expert judgment and received fit empirical evidence with partial credit model (PCM) based on polytomous data of four.
2. All question items in inquiry ability assessment test instrument were in good criteria because the difficulty level was in the range of -2,0 to 2,0. The highest test difficulty level was items of planning ability aspect, followed by implementation and reporting investigation results.
3. Reliability of inquiry ability assessment test instrument met the requirements, was in high category (reliability coefficient = 0,90).

Suggestions

1. Teachers should implement this inquiry ability performance assessment instrument in their schools to discover the progress of students' inquiry ability and use the results as assessment for learning for teachers.
2. Teachers should be facilitated with training in arranging inquiry ability performance assessment instrument.

REFERENCES

Bambang Subali.(2009). *Pengembangan tes pengukur keterampilan proses sains pola divergen mata pelajaran biologi SMA.* Prosiding Seminar Nasional Biologi, Lingkungan dan Pembelajaran, jurdik Biologi, FMIPA UNY, 4 Juli 2009, 581-593.

Bass, J.E, Contant, T.L, & Carin, A.A. (2009). *Theacing science as inquiry. 11^ed^*. USA: Allyn&Bacon.

Bond, T.G., & Fox, C.M. (2007). *Applying the rasch model: Fundamental measurement in the human sciences* (*2nd ed.*). Mahwah: Lawrence Erlbaum Associates, Publishers.

Bryce, T.G.K, McCall, J., Mac Groger, J., Robertson, I.J. and Weston, R.A.J. (1990). *Techniques for assessing process skill in practical science: teacher's guide*. Oxford:Heinemann Educational Books.

Chiappetta, E.L. & Koballa, T.R (2010). *Science instruction in the middle and secondary schools 7rd* ed. Pearson Education.Inc.

Collete, A.T. & Chiappetta, E.L. (1994). *Science instruction in the middle and secondary schools 3rd*. New York: Macmillan Publishing Company.

Curin & Sund. (1989). T*eaching science through discovery. Six Edition*. Merrill Publishing Company.

Depdiknas. (2003). *Standar kompetensi mata pelajaran fisika SMA/MA*. Jakarta : Puskur Balitbang Depdiknas.

Glencoe Science. *Performance assessment in the classroom*. Glencoe McGwaw-Hill.New York

Hambleton, R.K. & Swaminathan, H. (1985). *Item response theory*. Boston: KluwerNijhoff

Publishing.

Masters, G.N. (1999). *Partial credit model*. Dalam G.N. Masters & J.P. Kevees (Eds.), Advances in measurement in educational reseach and assessment. Amsterdam:Pergamon, An imprint of Elsevier Science.

Moh.Amien (1987). *Mengajarkan IPA dengan menggunakan metode discovery dan inquiry*. Departemen Pendidikan dan Kebudayaan. Jakarta

Moyer, R.H, Hackett, J.K, & Everett, S.A. (2007). *Teaching science as investigations modeling inquiry through learning cycle lessons*. New Jersey: Pearson, Merrill Prentice Hall.

Muraki, E. & Bock, R.D. (1998). *Parscale: IRT item analysis and test scoring for rating scale data.* Chicago: Scientific Softwere international, Inc.

Rezba, R.J, Sparague, C.S., Fiel, R.L, Fuck, H.J., Okey, J.R., & Jaus, H.H. (1995). Learning and assessing science process skill. 3^{-rd} ed. Iowa: Kendall/Hunt Publishing Company.

Ruiz-Primo, M.A. & Shavelson, R.J. (1996). *Rhetoric and Reality in science performance assessment: an update*. Journal of reseach and science teaching, Vol.33

Shavelson, R.J, Baxter, G.P., and Pine, J. (1992). *Performance assessment political rhetoric and measurement reality*. Research news & comment: Educational researcher.

Sumadi Suryabrata. (2000). *Pengembangan alat ukur psikologi*. Yogyakarta: Andi Ofset.

Supahar, Djemari Mardapi, and Zuhdan Kun Prasetyo. (2013). *Pengembangan tes untuk menilai kinerja kemampuan inkuiri mata pelajaran Fisika di SMA DIY*. Prosiding: Seminar Nasional Penelitian, Pendidikan, dan Penerapannya. Yogyakarta, FMIPA UNY,18 Mei 2013.

Wenning, C.J. (2005). *Implementing inquiry-based instruction in the science classroom*: A new model for solving the improvement-of-practice problem.*J. Phys. Tchr. Educ. Online 2*(4), May 2005.

_________, C.J. (2005). Levels of inquiry: Hierarchies of pedagogical practices and inquiry processes. *JPTEO, J. Phys. Tchr. Educ. Online 2*(3), February 2005.